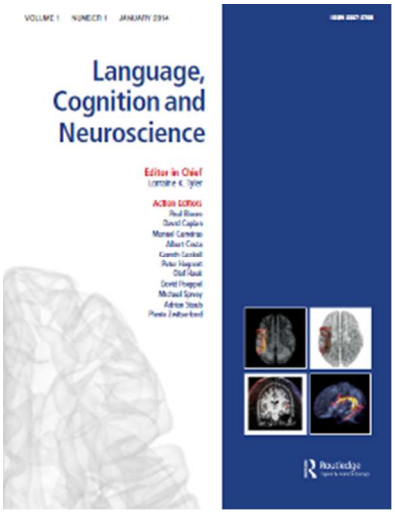
Yue J, Alter K, Howard D, Bastiaanse R. Early access to lexical-level phonological representations of Mandarin word-forms: Evidence from auditory N1 habituation.

***Language, Cognition and Neuroscience* (2017)**

DOI: https://doi.org/10.1080/23273798.2017.1290261



Early access to lexical-level phonological representations of Mandarin word-forms: Evidence from auditory N1 habituation

Journal:	*Language, Cognition and Neuroscience*
Manuscript ID	PLCP-2016-OP-9973.R3
Manuscript Type:	Original Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	YUE, JINXING; Harbin Institute of Technology, School of Foreign Languages Alter, Kai; Newcastle University, Howard, David; University of Newcastle-upon-Tyne, School of Education, Communication and Language Sciences Bastiaanse, Roelien; University Groningen, Linguistics;
Keywords:	short-term habituation, spoken word, auditory N1, Mandarin Chinese

Early access to lexical-level phonological representations of Mandarin word-forms: Evidence from auditory N1 habituation

Jinxing Yue [1,2*] Kai Alter[3,4] David Howard[5] Roelien Bastiaanse[6]

[1] International Doctorate in Experimental Approaches to Language And the Brain (IDEALAB), Newcastle University, University of Groningen, University of Potsdam, University of Trento, and Macquarie University

[2] School of Foreign Languages, Harbin Institute of Technology, Harbin, 150001, China

[3] Institute of Neuroscience, Newcastle University, Newcastle upon Tyne, NE2 4 HH, UK

[4] Faculty of Linguistics, Philology and Phonetics, University of Oxford, Oxford, OX1 2HG, UK

[5] School of Education, Communication and Language Sciences, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK

[6] Center for Language and Cognition Groningen (CLCG), University of Groningen, Groningen, 9700 AS, The Netherlands

Correspondence should be addressed to:

Jinxing Yue (岳金星)

Address: School of Foreign Languages, Harbin Institute of Technology, Harbin, China West Dazhi Street 92, 150001, Harbin, P. R. China

Email: yuejinxing@hit.edu.cn

This study was financially supported by Erasmus Mundus Joint Doctorate Fellowship issued by the European Commission [2012-1713/001-001-EMII EMJD], a grant for junior researchers by the National Social Science Fund of China [16CYY024] and a grant for Humanities and Social Sciences by Ministry of Education of P. R. China [12YJCZH262].

Word count: 8116

Abstract

An auditory habituation design was used to investigate whether lexical-level phonological representations in the brain can be rapidly accessed after the onset of a spoken word. We studied the N1 component of the auditory event-related electrical potential (ERP), and measured the amplitude decrement of N1 associated with the repetition of a monosyllabic tone word and an acoustically similar pseudo-word in Mandarin Chinese. Effects related to the contrastive onset consonants were controlled for by introducing two control words. We show that repeated pseudo-words consistently elicit greater amplitude decrements in N1 than real words. Furthermore, this lexicality effect is free from sensory fatigue or rapid learning of the pseudo-word. These results suggest that a lexical-level phonological representation of a spoken word can be accessed as early as 110 ms after the onset of the word-form.

Keywords: auditory N1; short-term habituation; spoken word; Mandarin Chinese

Introduction

Recognising a spoken word-form relies on rapid and accurate mapping of speech sounds onto pre-stored mental representations of phonological knowledge. It is widely accepted that the structure of phonological representations can be roughly divided into two levels: a lexical level and a sublexical level (Dahan & Magnuson, 2006). At the sublexical level, phonological knowledge is assumed to be stored as discrete units representing phonemes and non-lexical syllables. At the lexical level, a whole-word phonological representation is encoded as lexicalised connections of sublexical units. The time course of access to lexical and sublexical representations of phonology in processing spoken words has intrigued psycholinguists, and this interest has been recently enhanced by the application of neurophysiological recording techniques that provide high temporal resolution, such as electroencephalography (EEG) and magnetoencephalography (MEG).

A number of studies have reported that, given sufficient contextual information, the perception of some sublexical features (e.g., phonemes and syllables) can rapidly activate neural responses related to lexical processing, even before attention sets in (Pulvermüller et al., 2001; Shtyrov & Pulvermüller, 2002; Alexandrov, Boricheva, Pulvermüller, & Shtyrov, 2011; Shtyrov, Nikulin, & Pulvermüller, 2010; Shtyrov, 2011; Shtyrov, Kimppa, Pulvermüller, & Kujala, 2011; MacGregor, Pulvermüller, van Casteren, & Shtyrov, 2012; Gu et al., 2012; Shtyrov, Butorina, Nikolaeva, & Stroganova, 2014; Baart & Samuel, 2015). The basic rationale of these studies is that when a lexical-level representation is accessed in response to

a spoken word, the neural activation related to word-forms should be dependent on their lexical attributes (e.g., lexical status: real vs. pseudo-words) rather than on sublexical phonetic features.

Many of these studies used mismatch responses, such as Mismatch Negativities (MMNs) to index access of lexical-level representations. MMN, as elicited using an oddball paradigm, is a characteristic electroencephalographic response to *deviant* auditory stimuli (with low recurrence probabilities) interspersed in a sequence of more frequently presented *standard* stimuli (Näätänen, Gaillard, & Mäntysalo, 1978; Näätänen, 1995). In a seminal study, Pulvermüller and colleagues found that deviant Finnish words elicited more prominent MMNs than deviant pseudo-words in native speakers (Pulvermüller et al., 2001). The peak latency of the MMN effect was about 150 ms after the onset of the last syllable where a deviant stimulus became different from a frequently presented standard stimulus (e.g., Finnish deviant *lak<u>ki</u>* vs. standard *lak<u>ke</u>*). However, non-Finnish speakers did not exhibit any difference in MMN responses. Since pseudo-words lack lexical entries in the mental lexicon, these findings indicate fast access to lexical-specific long-term memory traces, represented in the cortical network.

More recently, an MEG study with English materials directly compared the neural responses to carefully matched real and pseudo-words that were delivered passively, rather than by an oddball paradigm (MacGregor et al., 2012). The researchers demonstrated distinctive, ultra-early response patterns with differences between real and pseudo-words, approximately 50-80 ms after the onset of word-final consonants, which were also the

recognition points after which the lexical status of a word-form could be decided (e.g., *cut* and **cuck*, where the asterisk indicates a pseudo-word). Their findings suggest that the human brain can access the neural circuitries encoding word-level memory at an ultra-rapid time scale.

Although the above-mentioned studies have provided some information about the capacity and latency of the human brain in the recognition of auditory signals as lexical input, most of these studies chose to observe the neural responses (associated with lexical access) from the onset of word-final consonants or syllables. This means that participants in those studies had already heard a great proportion of a word before receiving the final and essential part. However, the matter of when a lexical representation of phonology can be accessed after the *onset* of a word is less well studied.

Mandarin Chinese is a language model that allows us to deepen the investigation of the time course of lexical access. Mandarin (*Putonghua*) is characterised by its simple syllabic structure and its lexical tones. A syllable in Mandarin consists of an onset, a rime and a lexical tone. The onset and rime are segmental cues referring to initial consonants and vowels (a syllable nucleus), respectively. Lexical tones are pitch patterns, types of suprasegmental cues, overlaid on each syllable, and related to the fundamental frequency (f0). There are four lexical tones, namely Tone 1, Tone 2, Tone 3, and Tone 4. A syllable plays a crucial role in modern Chinese, because each syllable corresponds to at least one word or morpheme. Moreover, monosyllabic words are the word type with the highest frequency of use in modern Chinese (Wang et al., 1986). The meaning of a monosyllabic word is not only decided by its

segmental pattern, but also by its lexical tones. For example, a segment /ma/ has a completely different meaning when it carries contrastive lexical tones. That is, /ma1/ (妈) means "mother", /ma2/ (麻) means "hemp", /ma3/(马) means "horse", and /ma4/ (骂) means "curse" (Lee, 2007). In addition, there is a wide consensus in the literature that, like Indo-European languages, the mental representation of phonological knowledge in Mandarin Chinese also adheres to a lexical-sublexical structure (e.g., Ye & Connine, 1999; Zhao, Guo, Zhou, & Shu, 2011; Malins & Joanisse, 2012a). Specifically, representations of rimes, onsets, and tones are encoded as sublexical representations, while the lexicalised combinations of these units are represented separately at the lexical level.

By measuring event-related electrical potentials (ERPs) using scalp EEG, researchers have found that the perception of the acoustic cues of Mandarin words can be rapidly influenced by their lexical-level representations, in a top-down way (e.g., Zhao et al., 2011; Huang, Yang, Zhang, & Guo, 2014), suggesting that representations of lexical phonology can be accessed quickly after the onset of a spoken word. For example, Zhao and colleagues presented picture–word pairs to Mandarin speakers. The name of the object was shown in the picture and the paired word had phonetic mismatches. They found that participants had similar N400 responses to those words triggered partially syllabic mismatches in tones (/bi2/ "nose" vs. /bi3/ "pen"), onsets (/bi2/ vs. /li2/ "pear"), and rimes (/bi2/ vs. /bo2/ "neck"), in terms of both the N400's amplitudes and latencies (Zhao et al., 2011). The N400, a negative-gong ERP component, peaking around 400 ms post-stimulus onset, has been functionally associated with semantic processing and/or access to long-term semantic

memory (Kutas & Hillyard, 1984; Kutas & Federmeier, 2000; Kutas & Federmeier, 2011). Considering that the onsets occur temporally earlier than rimes and lexical tones, the results of Zhao and colleagues suggest that lexical access with tone-word stimuli relies on analysis of holistic phonological patterns in the N400 time window, approximately 300-500 ms after the onset of target auditory words (see Schirmer, Tang, Penney, Gunter, & Chen, 2005; Hu, Gao, Ma, & Yao, 2012 for similar results with Chinese languages; but Malins & Joanisse, 2012b for data demonstrating an incremental, instead of holistic processing pattern).

In contrast to these long latencies identified in the N400 time window, there are ERP effects that can be linked to the access of lexical-level phonology in much earlier time windows (e.g., Huang et al., 2014; Yue, Bastiaanse, Alter, 2014). For example, a recent study from our lab, using an oddball paradigm, showed that the MMN responses elicited by a monosyllabic Mandarin pseudo-word became word-like in the final phase of passive perceptual training (Yue et al., 2014). This effect was identified in two time windows, namely an ultra-early N1 time window (60–110 ms) and a typical MMN time window (198–238 ms). Furthermore, this changing pattern was found in the left scalp electrodes. Since MMN responses have been associated with long-term memory traces for words (Pulvermüller et al., 2001; Shtyrov & Pulvermüller, 2002), these results have been explained as the rapid learning of the pseudo-word (Shtyrov et al., 2010, Shtyrov, 2011; see Shtyrov, 2012 for a review). However, when the mismatch responses were analysed using the data from the entire training session, prominent responses in the N1 time window disappeared (see Shtyrov, 2011 for a similar observation). These findings have two important implications. Firstly, for

monosyllabic Mandarin words, lexical-level phonological representations may be accessed in the N1 time window, as early as 100 ms post-stimulus onset (Näätänen & Picton, 1987). However, the N1 is known to be very sensitive to the sensory features of acoustic input (Näätänen & Picton, 1987), and thus it is not easy to investigate processing at a lexical level within this time range. Secondly, the experimental paradigms used in previous studies, such as the oddball paradigm may not to be suitable for investigating lexical access in very early time windows prior to a standard MMN at 150-250 ms (Pulvermüller & Shtyrov, 2006).

In the present study, to address these limitations, we take advantage of a phenomenon known as "auditory habituation" to further examine whether the existence of lexical-level phonological representations induces differential brain responses to monosyllabic real and pseudo-Mandarin words in the auditory N1 time window. Habituation, in general, refers to a decrement in responsiveness to repeated stimulation, which is not due to sensory adaptation or fatigue[1] (Harris, 1943; Sharpless & Jasper, 1956; Thompson & Spencer, 1966; Rankin et al., 2009). Decreased responses typically recover after a stimulus-free period (Pinsker, Kupfermann, Castellucci, & Kandel, 1970). Thus, habituation of responses is considered to be the result of short-term neural plasticity in the respective neural network activated by stimulation (see Kandel, 2001 for a review). The proposed neural mechanism behind this phenomenon is that repeated stimulation (usually innocuous[2]), reduces the synaptic efficacy and connective strength between those responding to input and the neurons receiving output, leading to decrements in behavioural and electrophysiological responses (Castellucci, Pinsker, Kupfermann, & Kandel, 1970; Castellucci, & Kandel, 1974).

In the auditory domain, habituation has been reliably measured by the decreasing amplitude of N1, or its MEG event-related field (ERF) homologue N1m, typically recorded around the vertex[3] (e.g., Fruhstorfer, Soveri, & Järvilehto, 1970; Rust, 1977; Rosburg, Zimmerer, & Huonker, 2010; Wang & Knösche, 2013). This has been quantified by the decrease in amplitude of the N1 responses to each repeated stimulus in a stimulation train, as compared to the N1 amplitude to the first stimulus in the train. The interval between each stimulus presentation (i.e., the within-train inter stimulus interval, ISI) is usually constant and shorter than 10 s (e.g., 1 s, 3 s, and 10 s used in Budd, Barry, Gordon, Rennie, & Michie, 1998; 0.6 s, 1.2 s, and 1.8 s by Rosburg et al., 2010).

This kind of habituation is called "short-term habituation"[4], because decreasing N1 responses can recover after 4 -10 s of stimulation-free time (Rosburg, Haueisen, & Sauer, 2002; Rosburg et al., 2010; Wang & Knösche, 2013; Muenssinger et al., 2013). The decrease in response size seen with short-term habituation is typically a negative exponential function of the number of stimulus presentations (Rankin et al., 2009). However, auditory habituation studies have reported inconsistent patterns of response decrements even when using similar experimental designs. Some studies found a classical, gradual decrease as concluded by the short-term habituation literature (Fruhstorfer et al., 1970; Woods & Elmasian, 1986; Rosburg et al., 2002). Other studies identified unchanged decrements with the second subsequent presentations of a stimulus, described as a straightforward decrement pattern (e.g., Budd et al., 1998; Rosburg et al., 2006, 2010). Although some researchers have argued that the latter pattern is the consequence of neural refractoriness[5] instead of habituation (Ritter, Vaughan, &

Costa, 1968; Budd et al., 1998; Rosburg et al., 2010), we still use the term "habituation" to refer to the general response decrement, as the difference between the two decreasing patterns in the auditory domain has not yet been clearly identified.

Changes in N1 latency have also been previously reported, although these findings are much less consistent than those for the N1 amplitude. While a number of studies reported no prominent change in N1 latency (Rosburg, Haueisen, & Kreitschmann-Andermahr, 2004; Sörös, Michael, Tollkötter, & Pfleiderer, 2006; Sörös, Teismann, Manemann, & Lütkenhöner, 2009), other studies documented shorter latencies with repeated stimulation (e.g., Budd et al., 1998; Rosburg, 2004; Rosburg et al., 2010; Rosburg et al., 2006). It should be noted that one study has recorded *intra*cranial N1 with auditory stimuli in human subjects, and found no decrease in N1 latency with repeated stimulation (Rosburg et al., 2004).

In the present study we explore whether lexical-level phonological representations can be accessed in an N1 time window, during tone-word perception. A habituation design is an optimal tool for addressing this aim, for a number of reasons. Firstly, auditory short-term habituation is typically manifested by detectable neural responses in the N1 time window in which we are interested (Fruhstorfer et al., 1970; Rosburg et al., 2010).

Secondly, the auditory habituation of N1 has been shown to allow differentiation between neural representations at high and low levels corresponding to different types of stimulus (e.g., speech sounds at a "high" level versus pure tones at a "low" level). For example, one early study discovered that when the stimulus onset asynchrony (SOA) between two repeated stimuli was 1000 ms, the N1 amplitudes to speech sounds showed greater

decrease than to pure tones (Woods & Elmasian, 1986). In another study, with MEG, Teismann and colleagues (2004) found that the auditory N1 responses to repeated speech sounds had lower attenuation in the left hemisphere than in the right hemisphere, whereas the response decrements in acoustically matched tone stimuli were equivalent in both hemispheres. Furthermore, a few studies with functional imaging techniques have even shown that auditory habituation can be associated with phonological mismatch detection outside the focus of attention (Zevin & McCandliss, 2005; Joanisse, Zevin, & McCandliss, 2007).

Thirdly, habituation is induced with stimulus repetition rather than identification of acoustic/phonetic contrasts. Thus, it is not necessary to present phonologically similar real- and pseudo-words that contrast only in one phonetic feature in the same stimulation train or experimental block as would be required in an oddball paradigm (e.g., /peng3/ versus */teng3/ in Yue et al., 2014). As a result, a habituation design may unravel lexical-level processing in a more straightforward manner and decrease the risk of being confounded by acoustic change detection induced by contrastive phonemes, as can be an issue with oddball experiments (Näätänen et al., 1997; Dehaene-Lambertz, 1997). Moreover, a habituation-based design may be a way to avoid the rapid learning of a repeatedly presented pseudo-word. This is because short-term habituation only requires the repetition of one type of stimulus, and thus a learning context in which phonologically similar words and pseudo-words are presented in the same experimental block does not exist (e.g. Shtyrov et al., 2010; Shtyrov, 2011; Yue et al., 2014; Kimppa et al., 2015).

In this study, we compare the decrements in auditory N1 responses, as measured by EEG, elicited by a meaningful word /ma1/ and a novel word-form */na1/. We expect to observe differential short-term habituation patterns of the N1 to real and pseudo tonal word-forms, because auditory N1 habituation differs with stimuli accessing neural representation in different levels (e.g., speech sounds versus pure tone). The pseudo-word used in this study is tone-manipulated, meaning that it is synthesised by combining a existent segment pattern (/na/) with a lexical tone (Tone1) to form a novel segment-tone pattern (*/na1/) that is not meaningful in Mandarin Chinese. The merit of this type of pseudo-word is that it has legal segmental and tonal representations at a sublexical-level (e.g., /na/ + Tone2 (拿) means "take with hands"), without lexical-level entries in the mental lexicon. Therefore, any response difference between a tone-manipulated pseudo-word and an acoustically matched real word is likely induced by the lack of a lexical-level phonological representation of the pseudo-word, rather than by low-level phonological processing difficulties (MacGregor et al., 2012; Yue et al., 2014). To rule out the possibility that any observed effects are induced by the contrast between the onset consonants (/m/ versus /n/), we also present two real words with the same onset contrast (/mi2/ versus /ni2/) as control conditions.

Method

Participants

Twenty native speakers of Mandarin Chinese (i.e., *Putonghua*) were recruited for this study.

All participants reported normal hearing, no history of neurological disorders, and no abuse of alcohol or drugs in a pre-experiment questionnaire (male: 8; *mean age* = 29, *SD* = 4, range = 22 - 39). All had acquired Mandarin Chinese from early childhood without exposure to dialects in which the pseudo word-form /na1/ is meaningful (as it is in *Lanzhou* and *Xuzhou* dialects). All participants were right-handed, as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971) with minor adaptations for Chinese culture (e.g., replacing "knife" with "chopsticks"). Written informed consent was given by participants prior to the experiment, according to Declaration of Helsinki under a procedure approved by the Ethical Committee of the Faculty of Arts, University of Groningen. Participants were paid 20 euros as compensation for their participation.

Stimuli

To observe the habituation of the N1, we utilised a short-term habituation design in which a stimulus was repeatedly presented in a stimulation train with a short and constant ISI (Teismann et al., 2004; Rosburg et al., 2010; Wang & Knösche, 2013). By repeating this stimulation train a certain number of times, reliable auditory habituation at each stimulation position can be obtained by comparing the averaged amplitudes of responses to the first stimulation with those to the subsequently presented stimuli in each position.

We constructed four stimuli to give two experimental conditions and two control conditions. In one experimental condition, the stimulus was a tonal pseudo word-form */na1/; the other experimental condition consisted of a real tone word /ma1/ ("mother", 妈). These

two word-forms carried identical rimes and tones with acoustically similar onsets, allowing us to compare the N1 habituation patterns in stimuli with and without lexical-level phonological representation. For the two control conditions, two real words, /ni2/ ("mud", 泥) and /mi2/ ("mystery", 迷) were used to control for potential habituation effects induced by the onset contrast (/n/ versus /m/). Table 1 presents a summary of the tonal word-form stimuli. The three real words are all frequently used in modern Mandarin Chinese, according to the *List of Frequently Used Words in Modern Chinese* (Language Affairs Revolution Council of China, 1988) [6].

(Insert Table 1 about here)

During the experiment, we presented each stimulus condition in a separate experimental block, meaning four blocks per experiment participant (Rosburg et al., 2010). Each block consisted of a total of 75 stimulation trains with an inter-train interval (ITI) of 5 s. Each stimulation train consisted of five presentations of the same target stimulus (denoted S1, S2, S3, S4, and S5) with inter-stimulus interval (ISI) of 500 ms (see Figure 1). Sixty of the 75 trains were presentations of the target stimulus for that condition. In addition, three other word-forms, without segmental or tonal overlaps with the target item, were presented as fillers (see Appendix for all foil words and pseudo-words). Five trains of each filler (15 in total) were presented, pseudo-randomly mixed with the 60 trains of target items, such that trains of the target item were not presented more than four times in a row.

The overall experimental paradigm for a subject consisted of four blocks in one of the

two orders: */na1/-/mi2/-/ma1/-/ni2/, or /mi2/-*/na1/-/ni2/-/ma1/. This arrangement was used to avoid too many repetitions of the same lexical tone and rime vowel in successive blocks. Additionally, in order to avoid potential rapid-learning effects on the perception of a pseudo-word, under a context in which acoustically similar real words were also presented (Shtyrov et al., 2010; Shtyrov, 2011; Yue et al., 2014), the experimental block for the pseudo-word */na1/ was always presented prior to the block for the real word /ma1/. Participants were randomly assigned to one presentation order, and the two orders were counterbalanced between participants. There was a 1.5-minute stimulus-free period between the first and the second blocks, and between the third and the fourth blocks. A 2-minute break was set between the second and the third blocks. Control of stimulus delivery and trigger delivery was implemented in E-prime software (Version 2.0, Psychology Software Tools Inc., Pittsburgh, PA, USA).

(Insert Figure1 about here)

To ensure that the measured responses reflected natural-word-related brain activities, we used spoken word-forms articulated by a well-trained female native speaker of Mandarin Chinese. Forty-five exemplars for each type of stimulus were produced and recorded in a sound-proof recording studio with Adobe Audition (version 3.0), through a Shure (SM27) microphone, a Dateq (BCS25) audiomixer, and an Adirol (UA 25) USB audio capture device digitising at a sampling rate of 44.1 kHz (bit-depth, 16 bit). One of the exemplars for each condition was selected. Criteria for selection included the following: the candidate word-form

should be neatly articulated with an approximate duration of 500 ms before normalization; preferred candidates with the same lexical tone (e.g., /ni2/ and /mi2/) should contain similar pitch contours and pitch heights.

The four chosen candidates were normalised to the same duration (500 ms). After this procedure, we replaced the fundamental frequency of candidate /ma1/ with the one of candidate */na1/ (f0 onset: 246 Hz; offset: 247 Hz; minimum: 237 Hz; maximum: 247 Hz); the f0 of candidate /mi2/ was overlaid by that of candidate /ni2/ (f0 onset: 197 Hz; offset: 280 Hz; minimum: 197 Hz; maximum: 280 Hz). Lastly, all four candidates were adjusted to the same average intensity (75 dB). To ensure that the control words met our experimental requirements, we calculated the average f2 values of the onset consonants of the four types of stimuli. F2 has been considered as a key acoustic measure for distinguishing between /m/ and /n/ in Chinese (Li, Zhang, & Huang, 2009). The f2 values for the /m/ in normalised /ma1/ and /mi2/ were 1409 Hz and 1583 Hz, and those for the /n/ in */na1/ and /ni2/ were 2506 Hz and 2179 Hz, respectively. These data confirmed that the control words were suitable for revealing effects induced by the onset contrast. All these steps were performed using Praat software (Boersma & Weenink, 2013). Figure 2 displays the oscillograms and spectrograms of the materials used.

(Insert Figure 2 about here)

EEG data acquisition

Participants were tested individually in a sound-attenuated room with dim lighting, and seated comfortably in front of a video monitor at a distance of about 1.2 m. To control for the reported impact of selective attention to auditory stimuli on habituation (Öhman & Lader, 1972), participants were instructed to watch a silent cartoon movie and to remember the contents, in order to achieve a good performance in a post-experimental comprehension test. In the test, participants were asked to judge whether 12 statements matched what they saw in the movie by choosing either "Yes", "No", or "Cannot remember". They were also asked to ignore any auditory interference (actually the stimuli) delivered binaurally via headphones at 55 dB SPL during the movie.

The EEG data were collected at a sampling rate of 500 Hz with an ANT acquisition system (Advanced Neuro Technology, ANT, Enschede, The Netherlands). 64-channel WaveGuard caps (ANT, extended international 10-20 system) with shielded wires to decrease the influence of environmental electrical fields were used to record EEG signals from the scalp, as well as from the two mastoids. To record artefacts caused by ocular movements and blinks, two electrodes were placed on the outer canthi of both eyes for horizontal movement detection, and two electrodes were situated on the left infraorbital and supraorbital ridges to monitor vertical activities. The ground electrode was AFz. An average reference was used for

online EEG acquisition. The impedance of each electrode was kept below 5 kΩ. The EEG recording protocol took about 100 minutes to complete.

EEG data processing and analysis

For offline data processing, the EEG recordings were first bandpass filtered at 1-30 Hz (24 dB/Oct). The data were then re-referenced to the average of the two mastoids. Epochs were segmented separately for the five presentations of each stimulus type, with a window of 600 ms around the stimulus onset (-100 ms before to 500 ms after). The baseline was corrected to the pre-stimulus period. Trials contaminated with artefacts (> 100 μV voltage variation) were rejected. ERPs were obtained by averaging the remaining epochs (an average of 36 trials per presentation per condition).

The N1 peaks were detected at each electrode on a per participant basis. The N1 peak detection for local maxima was confined to a time window between 80 and 170 ms, defined *a priori* to cover all latency ranges for all N1 components (Pang & Taylor, 2000; Mulert et al., 2005; Bell, Dentale, Buchner, & Mayr, 2010). The decrement in the N1's peak amplitudes caused by habituation was indexed by habituation coefficient *R*, calculated as $R = N1_{Sn} / N1_{S1}$ *(n = 2, 3, 4, 5)*, in which $N1_{Sn}$ refers to the N1 peak amplitude with stimulus S_n. The lower the value of *R*, the more the N1 response was habituated. This normalisation procedure was performed per participant from the data collected at nine electrodes around the vertex (FC3, FCz, FC4, C3, Cz, C4, CP3, CPz, CP4). The reason we analysed these electrodes is that N1 habituation at the vertex has been shown to be the most prominent (Fruhstorfer et al., 1970;

Fruhstorfer, 1971) and the electrical potential changes recorded at these positions have typically been used to indicate neural activity associated with auditory/speech perception (e.g., Prescott, Connolly, & Gruzelier, 1984; Shtyrov et al., 2010). The nine electrodes were grouped into three regions of interest (ROIs), capturing the N1 decrement in the left- (FC3, C3, FC3), midline- (FCz, Cz, CPz), and right- (FC4, C4, CP4) scalp regions around the vertex (Cz). *R* data of the electrodes were averaged for each ROI. To decrease the noise in the ROI analysis, *R* values greater or smaller than 2 *SD* were excluded. Participants who had no remaining data after the trimming procedure for any one of the ROIs were treated as outliers and disregarded [7].

Statistical analyses were performed to examine the lexicality effect in the N1 habituation. The analysis began with a 4×2×2×3 repeated measures ANOVA with four factors: Repetition (S2-S5), Condition (experiment versus control), Onset (/m/ versus /n/), and Laterality (left, midline, and right hemispheric ROIs). The lexicality effect was tested by identifying any interactions between the factors Condition and Onset. Once the lexicality effect had been assessed, *post-hoc* tests were carried out to examine the source of interaction. If the source could be located, planned comparisons would be performed to test whether the lexicality effect was induced by the contrastive responses between the two experimental conditions.

To make sure that the lexicality effect was habituation-based, another set of *post-hoc* analyses were performed to examine the role of sensory fatigue and/or rapid-word learning that might occur during the experiment due to repeated presentation of the same auditory stimulus. We calculated the average *R* values by dividing all trials into two phases: the first

30 and last 30 trials which represented the amplitude decrement of the N1 in the "early" and "late" experimental phases, respectively. Since this analysis was based on the existence of a lexicality effect between the N1 habituation to /ma1/ and */na1/, a repeated measures ANOVA was conducted with four factors: Phase (early versus late experimental phases), Repetition (as above), Laterality (as above), and Lexicality (/ma/ versus */na1/) to test for a main effect of Phase and its interactions with Lexicality. If the N1 decrements were influenced by sensory fatigue, we would expect a main effect of Phase. If the repetition of the pseudo-word induced rapid word learning, the N1 habituation to this pseudo-word should be word-like in the late phase of experiment but not in the early phase, and thus an interaction between Phase and Lexicality would be expected.

We also conducted analysis of N1 peak latencies, in order to examine whether the temporal patterns of habituated N1 responses were different between real and pseudo word-forms with lexical tones. The N1 peak latencies were averaged over the three ROIs separately. A four-way ANOVA was run with four factors: Presentation (S1 to S5), Laterality (as above), Condition (as above), and Onset (as above), to test for main effects of each of the factors, and an interaction between Condition and Onset.

For each ANOVA, Greenhouse-Geisser correction was performed when the sphericity assumption was violated. The significance level (alpha) for type I error was set at $p< 0.05$. Bonferroni corrections were applied to correct p values in multiple t-tests. All statistical tests were performed in SPSS (Version 20, IBM Corporation, Armonk, U.S.).

Results

In the cartoon comprehension test, only an answer of "Yes" was counted as correct. The average percentage of correct answers was 90% (*SD* = 10%). The high rate of correctness indicates that participants' attention was well focused on the movie.

Visual inspection

Visual inspection of the ERP wave forms and the topographic maps showed evidence of decreasing N1 responses with stimulus repetition in all four conditions (Figures 3, 4). Therefore, the chosen auditory habitation paradigm fulfilled the purpose of eliciting short-term habituation in the auditory N1. The habituation coefficient *R* values for the right, midline, and left ROIs confirmed the results of the visual inspection (Table 2). The repeated word stimuli induced about 40% decrement in N1 peak amplitudes relative to the responses to the first stimulation (Table 2). There seemed to be no clear differences between the habituated N1 responses to each of the four types of stimuli, except in the right ROI, where there were greater decrements in the N1 peak amplitudes for the pseudo-word than for the three real words (bold columns in Table 2). Regarding the peak latencies of N1, these were centred around 110 ms for all five presentations (e.g., see Cz in Figure 3) and the average peak latencies in the three ROIs were in the range 101 – 124 (Table 3).

(Insert Figure 3 about here)

(Insert Figure 4 about here)

(Insert Table 2 about here)

(Insert Table 3 about here)

N1 decrement analysis

We identified an interaction between Condition × Onset × Laterality (F (2, 34) = 4.58, p = 0.03) from the ANOVA of the average R values. However, we found no interactions between Condition × Onset and Repetition (Repetition × Condition × Onset × Laterality: F (6, 102) = 0.47, p = 0.74; Repetition × Condition × Onset: F (3, 51) = 0.39, p = 0.70). These results suggest that the N1 amplitude decrement is influenced by the lexical status of stimuli in some ROIs, but the four repeated presentation positions share generally similar N1 decrement patterns, without systematic variation between different conditions.

Motivated by the three-way interaction, we conducted separate ANOVA tests in the left, midline and right ROIs, revealing a two-way interaction between Condition × Onset only in the right ROI (F(1, 17) = 4.61, p = 0.046). We performed two planned comparisons which showed that Onset was a main effect in the experimental conditions (*/na1/ versus /ma1/, F(1, 17) = 10.21, p = 0.005), but not in the control conditions (/ni2/ versus /mi2/, F(1, 17) = 0.06, p = 0.82), as shown in Figures 3, 5, and 6. Descriptive statistics showed that the average coefficient R in the right ROI for the pseudo-word was about 12% smaller than that of the experimental real word and the two control words (*/na1/: M = 0.47, SD = 0.27; /ma1/: M = 0.64, SD = 0.31; /ni2/: M = 0.56, SD = 0.28; /mi2/: M = 0.57, SD = 0.31). In addition, the latencies of the N1 responses in this region were around 110 ms post-stimulus onset, for all four stimulus types (range 104 - 122 ms). These results suggest that in line with our

expectation that the short-term habituation of N1 differs between real and pseudo word-forms.

(Insert Figure 5 about here)

(Insert Figure 6 about here)

We conducted *post-hoc* analysis to examine whether the N1 habituation was influenced by sensory fatigue and/or the rapid word-learning mechanism. A four-way ANOVA revealed only an interaction between Lexicality and Laterality (F (2, 34) = 4.09, p = 0.03). Further analyses showed that there was a main effect of Lexicality only in the right ROI (F (1, 17) = 5.38, p = 0.033), which was consistent with the results presented above. Furthermore, we found neither a main effect of Phase (F (1, 17) = 0.39, p = 0.54) nor an interaction with Lexicality (F (1, 17) = 3.48, p = 0.08). These results confirm that the lexicality effect we found in the right ROI was indeed caused by auditory habituation, and was free from sensory fatigue and rapid word learning.

Latency analyses

A four-way ANOVA only revealed a main effect of Onset (F(1, 17) = 60.91, p< 0.001). *A post-hoc t*-test revealed that stimuli with an onset of /n/ triggered earlier N1 peaks than those with an onset of /m/ with a difference of around 13 ms (/n/ latency M = 105 ms, /m/ latency M = 118 ms; /n/ versus /m/: t (539) = -17.44, p< 0.001). The main effects of the factors Presentation, Condition, and Laterality were not found to be significant (Presentation: F(4, 68) = 1.29, p = 0.28; Condition: F(1, 17) = 0.18, p = 0.69; Laterality: F(2, 34) = 0.64, p = 0.47). These results suggest that the N1 latencies did not vary with stimulus repetition.

Discussion

A number of studies have presented neurophysiological evidence that the human brain is able to rapidly access lexical-level representations of words upon the perception of sublexical features (Pulvermüller et al., 2001; MacGregor et al., 2012). However, most of these studies chose to observe lexical-level processing at the word-final time point (e.g., the onset of the last phoneme or syllable), at which a large proportion of a word-form has already been received. What is less well understood is the time course of access to phonological representation when measured from the word onset. To fill this gap, we recorded the EEG of 20 native Mandarin Chinese speakers during an auditory habituation experiment using tonal word-forms to look for evidence of access to lexical phonology in the N1 time window.

We observed the decrements of auditory N1 peak amplitudes in response to repeatedly presented spoken Mandarin word-forms. A lexicality effect was revealed by the difference between the N1 habituation to a pseudo word-form (*/na1/) and the habituation to an acoustically matched real word (/ma1/). To the best of our knowledge, this is the first study reporting evidence that the lexical status of a spoken word-form modulates the habituation of the N1, round 110 ms post stimulus onset.

Habituation of the N1

Using an auditory habituation design, we successfully elicited short-term habituation of the spoken-word-related N1 component with stimuli presented five times in a stimulation train,

separated by a constant ISI of 500 ms. The magnitude of the N1 amplitude decrement in our study is comparable to the results of one auditory habituation study with speech stimuli. Woods and Elmasian (1986) reported that the N1 amplitudes to speech stimuli at the vertex for the fourth presentation position were approximately 52% of the N1 responses to the initial stimulation. Similarly, in the current study, the average N1 peak amplitudes of the three real-word stimuli /ma1/, /ni2/, and /mi2/ in the midline ROI for the same presentation position are approximately 61%, 53%, and 59% of the N1 to the first stimulation S1 (Table 2). Furthermore, our data demonstrate that the peak latency of the N1 response to tonal word-forms does not vary with the repetition of stimuli. This temporal pattern is line with a number of previous studies using pure tones, including one that recorded intracranial N1 responses to auditory stimuli (Rosburg et al., 2004; Sörös et al., 2006; Sörös et al., 2009).

Unlike the gradual amplitude decrease pattern described in classical habituation (Thompson & Spencer, 1966; Rankin et al., 2009), our results show straightforward decrements in the N1 peak amplitudes. That is, there is no systematic difference in the N1 decrement between S2 and the later presentation positions. In other words, the decreasing N1 responses in the current study have already reached a minimum level with the first repetition (S2), whereas the habituation of the N1 in many studies has been found to be complete by the third stimulation (e.g., Woods & Elmasian, 1986; Rosburg et al., 2002). Our results seem to suggest that the auditory N1 decrement to tonal word-forms is induced by neural "refractoriness" rather than "habituation". However, we do not think that distinguishing between refractoriness and classical habituation models is helpful in explaining our results for

a number of reasons. Firstly, previous studies, using similar experimental designs to the current study, have reported either gradual or straightforward decrements in auditory-evoked responses (e.g., gradual decrease in Woods & Elmasian, 1986; straightforward decrease in Budd et al., 1998). Therefore, the distinction between neural refractoriness and habituation in auditory responses is still far from clear, especially when auditory responses are measured by EEG or MEG from the scalp. Secondly, from the viewpoint of neurobiology, short-term habituation has been attributed to a depletion of releasable vesicles containing neurotransmitters, as a result of successive stimulation (Castellucci et al., 1970; Castellucci & Kandel, 1974; Sara, Mozhayeva, Liu, & Kavalali, 2002). Such a physiological mechanism is very likely to be at least partially shared by refractoriness and habituation, and, thus, the two models of auditory response decrement are not necessarily mutually exclusive (Rosburg et al., 2010). All in all, we suggest that the habituation design allows us to capture the brain's electrical responsiveness to repeated word-forms, reflecting how the human cortex becomes habituated to the repeated input of both real and pseudo tonal word-forms.

Lexicality modulation of the tone-word-related N1 habituation

The most important finding of the current study is that native Mandarin speakers' responses to repeated word-forms in a N1 time window (104 - 122 ms) can be modulated by the lexical status of the stimulus. This means that word-forms with and without phonological representations at a lexical level elicit divergent N1 habituation patterns. By using five-pulse stimulation trains, we have shown that the repetition of a pseudo word-form elicits greater

decrements in N1 peak amplitudes than those induced by an acoustically similar real word over the right-hemispheric scalp region. Such N1 habituation pattern is consistent throughout the four repeated presentations of a stimulus (i.e., S2 – S5). Additionally, the result of a *post-hoc* analysis confirmed that the lexicality effect was based on habituation and was free from sensory fatigue.

We interpret this lexicality effect as evidence of rapid access to a lexical-level phonological representation during tone-word perception. The reasons are as follows. Firstly, we have carefully matched the physical features of the pseudo-word and the real word to ensure that the major representational difference between the two types of stimulus is at the lexical level. Specifically, the pseudo-word does not involve any illegal segmental or suprasegmental parts, but has no lexical-level representation; thus, the processing of such pseudo-words can only access the neural representations of phonology at a sublexical level (i.e., those for consonants, vowels, and lexical tones), rather than the neural circuitries encoding lexical phonology. In contrast, the processing of real words is able to access higher-order neural representations at a lexical-level, not just the low-level phonological units. Hence, the distinction in the N1 decrement between the real and pseudo forms of tonal words is very likely to be due to differences in their neural representations, namely the existence of a lexical-level phonological representation for real words but not for pseudo-words. Our representational account is in line with previous studies which reported that the N1 habituation to speech sounds was different from the habituation to pure tones (Woods & Elmasian, 1986; Teismann et al., 2004).

Secondly, although the pseudo-word and real-word stimuli have contrastive onsets (/n/ and /m/), a comparison with two control words (/ni2/ and /mi2/) indicated that the lexicality effect in N1 habituation cannot be attributed to the processing of acoustic or phonetic differences via sublexical representations. Specifically, the N1 amplitude decrements in the two words did not differ from each other. Neither did those between the real word in the experimental condition (/ma1/) and the two control words. In addition, since the pseudo and real word-forms in the experimental conditions have identical tone and rime parts, there is no reason to associate the lexicality effect with any processing at a sublexical level such as tone processing.

Thirdly, since the N1 time window is far too early for other aspects of lexical processing, we hold that the differential N1 habituation patterns between the real word and the pseudo-word are induced by processing within the phonological domain. We acknowledge that a real word and a meaningless pseudo-word also differ in terms of semantics (e.g., meaning). However, there is little evidence that brain responses in the N1 time window are sensitive to semantic manipulations. It is widely accepted that the N1 time window is typically associated with phonological processing, whereas the ERP correlates of semantic processing are usually much later than 150 ms, at 300-500 ms after the onset of a word (see Friederici, 2002 for a review of auditory word recognition in sentence contexts; see Pulvermüller, & Shtyrov, 2006 for a review of isolated word perception).

The underlying cognitive processes

The latencies of the N1 peak suggest that, for monosyllabic Mandarin words, access to lexical-level phonological representations can occur very early: around 110 ms after the onset of a spoken tonal word-form. This result is in line with our previous ERP study reporting lexical mismatch responses to monosyllabic Mandarin word-forms in 60-110 ms time window (Yue et al., 2014), which is earlier than the time windows reported for lexical-level processing, ranging between 200 ms to 500 ms post-stimulus onset in the ERP literature (e.g., Schirmer et al., 2005; Zhao et al., 2011; Huang et al., 2014). These results are consistent with studies in Indo-European languages, which show that the processing of a speech input can rapidly access lexical representation in the human brain as early as 50-100 ms (Pulvermüller et al., 2001; Shtyrov & Pulvermüller, 2002; Shtyrov et al., 2010; MacGregor et al., 2012). However, we should reiterate that studies finding evidence of rapid lexical access in Indo-European languages were based on the perception of a large proportion of the word-form. In contrast, the lexicality effects in the current study and our previous MMN study (Yue et al., 2014) were observed when acoustic input was very limited. That is, lexical-level phonology can be accessed following the input of very small number of segmental and suprasegmental cues.

Admittedly, an N1 time window, measured from a word's onset, is too short for word recognition and the perception of all sublexical components, such as rimes and tones (Wu & Shu, 2003; Lai & Zhang, 2008; Shuai & Gong, 2014). For example, Wu and Shu studied the isolation points of words, onsets, rimes, and lexical tones in monosyllabic Mandarin words.

They found that the reliable isolation of words, rimes, and lexical tones only occurred after 200 ms following the beginning of a naturally articulated word without any context. However, in a repetitive context as was used in the current study, it is likely that lexical entries can be activated before the perception of all sublexical features, since the phonological representation may have been pre-activated by the prior presentation of a repeated stimulus.

For real words, pre-activated lexical-level representations of phonology may have a lower access threshold (Gaskell & Marslen-Wilson, 2002; Huang et al., 2014), and thus, permit near-instant lexical access given very little acoustic input through processing at the sublexical level, via the interconnections between the two representational levels (Ye & Connine, 1999; Zhao et al., 2011; Malins & Joanisse, 2012a). In contrast, the pseudo-word stimuli can activate no lexical entries, but only the corresponding representations at the sublexical level. Therefore, the repetition of pseudo-words may only facilitate the access to the representation of sublexical features. As a result, such a representational difference is the likely reason that auditory N1 habituation to the real word diverges from that in the pseudo-word even before the auditory system receives all the segmental and suprasegmental cues of a word. However, we must acknowledge that, since the onset consonants in this study are sonorant in nature, the onsets are subject to subtle co-articulation, and thus may convey suprasegmental cues prior to the presentation of vowels (Howie, 1974; Lai & Zhang, 2008), which might contribute to the fast lexical access. This issue merits further studies to compare the N1 habituation in words with obstruent (e.g., /p/) and those with sonorant onsets.

Our account also applies to our previous discovery of the lexical effect in an N1 time

window (Yue et al., 2014). In that study, only two word-forms were repeatedly presented in one experimental block which may have induced early top-down processing due to the lower activation load of the lexical representation of the target word as compared to the processing of the repeated, unlearned pseudo-word. Therefore, it is reasonable to conclude that a repetitive context is crucial for early access of the lexical-level phonological representation of monosyllabic tone words.

There is one more interesting temporal issue of note in this study. That is, the specific time course of the lexicality effect was actually related to the phonetic features of the onsets. The analysis of N1 peak latencies showed that stimuli with an onset /n/ triggered slightly, but consistently, earlier N1 peaks than those with an onset /m/, suggesting that the latencies of N1 responses to word-forms are modulated by the phonetic features of the onset consonants. This kind of temporal-onset association shows that the N1 component is sensitive to the processing of sensory characteristics of stimuli (Näätänen & Picton, 1987). However, further studies are needed to pinpoint why decrements in the N1 amplitude were related the lexical status, whereas the peak latencies of the N1 were sensitive to the acoustic features of word onsets.

With respect to the spatial distribution of the lexicality effect, our results suggest that differential N1 habituation patterns between the real word and the pseudo-word are located in the right-hemispheric scalp region. That is, the N1 responses to repeated real words were more persistent than those to pseudo-words in signals recorded from electrodes situated over the right scalp, but no systematic difference between the two was found with electrodes at the

midline and left scalp regions. This finding contradicts our previous MMN study which reported that lexicality-related mismatch responses in an N1 time window were identified at the left-hemispheric scalp region (Yue et al., 2014).

Our tentative explanation for this issue is that, although the lexicality effects reported in the two studies are within similar time windows, the different experimental paradigms employed in the experiments may mean that distinctive auditory processes were involved. In the oddball paradigm (Yue et al., 2014), the processing of target word-forms (i.e., the deviants) in an N1 time window was based on auditory discrimination between the deviants and standards with contrastive onsets (Pulvermüller, & Shtyrov, 2006). In comparison, the way that a stimulus was presented in the auditory habituation design was extremely consistent, and the presentation of latter stimuli in a stimulation train was exactly the same as the first item in the train. Therefore, the underlying neural processes are very likely to be different between discrimination-based and repetition-based lexical access. Our explanation is consistent with previous studies which reported that the N1 and MMN can be associated with separated mechanisms due to different conditions of elicitation. For example, the auditory N1 must be elicited by the presentation of a stimulus, whereas the MMN can also be triggered when a deviant stimulus is omitted (Nordby, Hammerborg, Roth, & Hugdahl, 1993; Yabe et al., 1998). Furthermore, comatose patients exhibit higher sensitivity to auditory stimulation as indexed by N1 than by MMN (Fischer et al., 1999; Fischer, Morlet, & Giard, 2000).

Nevertheless, the results in the two studies are not necessarily mutually exclusive. Rather, they together imply that access of lexical-level representations of phonology is a complex

process, involving different neural mechanisms even in the very beginning stage of word perception. Our findings in the two studies also indicate that an auditory habituation design and an oddball paradigm may be selectively sensitive to different aspects of lexical-level processing during spoken-word perception, which deserves further comparative investigation.

Conclusion

Our study presents the first clear evidence that the existence of lexical-level representations could result in differential N1 habituation patterns between real and pseudo word-forms derived from Mandarin Chinese in native listeners. The results suggest that there is access to lexical-level phonological representations of spoken words as early as 110 ms after the onset of a word, when there is very limited acoustic input. Cross-linguistic studies are needed to test whether the observed lexicality effect of the auditory N1 habituation in repeated monosyllabic Mandarin materials also applies to other types of tonal words and/or words in Indo-European languages.

Acknowledgements

We thank Prof. Ben Maassen, Dr. Peng Wang, Dr.Jaap van der Spek, and Ms. Kelly Callahan for their help at different phases of this study. We also thank the reviewers for their insightful comments. This work was financially supported by Erasmus Mundus Joint Doctorate Fellowship issued by the European Commission [2012-1713/001-001-EMII EMJD], a grant

by the National Social Science Fund of China [16CYY024] and a grant for Humanities and Social Sciences by the Ministry of Education of P. R. China [12YJCZH262].

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes

1. Since the purpose of this study was to compare the pattern of response decrement with real words and with pseudo-words, instead of characterising the habituation phenomenon with word-related ERPs, we did not include any dishabituators in the current design. Dishabituators are usually used to check whether the response suppressions observed in the study reach the working criteria for habituation (Rankin et al., 2009). Nevertheless, using the current design which has been repeatedly used in many previous studies, we believe that the response decrement in N1 responses can reflect auditory habituation with repeated presentation of Mandarin word-forms.

2. Harmful stimulation usually leads to enhanced responses with stimulation repetition, called sensitisation (Kandel, 2001).

3. We used N1 to refer to both ERP N1 and ERF N1m.

4. The other type of habituation is auditory long-term habituation, which has also been observed by comparing N1s between experimental blocks or with an ISI longer than 10 s.

However, since there is still no clear correlation between any underlying sensory processing and long-term habituation (e.g., Öhman & Lader, 1977; Woods & Elmasian, 1986), our study only focused on short-term habituation. Therefore, short-term habituation and habituation are used interchangeably through this paper.

5. Neural refractoriness, or the refractory period, refers to the brief period when a neuron is not able to respond to another stimulus, immediately after the production of an action potential elicited by the previous stimulus (Mildner, 2008).

6. We did not match the exact word frequencies of the three real words, since the three words are not only used as words, but they are also morphemes to formulate polysyllabic words. Therefore, the word frequency may not reflect the true familiarity of a monosyllabic word.

7. Two participants were excluded from statistical analyses according to these criteria.

References

Alexandrov, A.A., Boricheva, D.O., Pulvermüller, F., & Shtyrov, Y. (2011). Strength of word-specific neural memory traces assessed electrophysiologically. *PLoS ONE*, 6: e22999. doi: 10.1371/journal.pone.0022999

Baart, M., & Samuel, A.G. (2015). Early processing of auditory lexical predictions revealed by ERPs. *Neuroscience Letters, 585*, 98–102. doi: 10.1016/j.neulet.2014.11.044

Bell, R., Dentale, S., Buchner, A., & Mayr, S. (2010). ERP correlates of the irrelevant sound effect. *Psychophysiology, 47*, 1182–1191. doi: 10.1111/j.1469-8986.2010.01029.x

Boersma, P., & Weenink, D. (2013). Praat: Doing phonetics by computer (Version 5.3.39).

Institute of Phonetic Sciences of the University of Amsterdam. <http://www.praat.org>.

Budd, T.W., Barry, R.J., Gordon, E., Rennie, C., & Michie, P.T. (1998). Decrement of the N1 auditory event-related potential with stimulus repetition: Habituation vs. refractoriness. *International Journal of Psychophysiology, 31*, 51–68. doi: 10.1016/S0167-8760(98)00040-3

Castellucci, V.F., Kandel, E.R. (1974). A quantal analysis of the synaptic depression underlying habituation of the gill-withdrawal reflex in Aplysia. *Proceedings of the National Academy of Sciences of the United States of America, 71*, 5004–5008. doi: 10.1073/pnas.71.12.5004

Castellucci, V., Pinsker, H., Kupfermann, I., Kandel, E.R. (1970). Neuronal mechanisms of Habituation and dishabituation of the gill-withdrawal reflex in Aplysia. *Science, 167*, 1745–1748. doi: 10.1126/science.167.3926.1745

Dahan, D., & Magnuson, J.S. (2006). Spoken-word recognition. In M. J. Traxler& M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (pp. 249–283). Amsterdam, The Netherlands: Academic Press.

Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *Neuroreport, 8*, 919–924. doi: 10.1097/00001756-199703030-00021

Fischer, C., Morlet, D., &Giard, M. H. (2000). Mismatch Negativity and N100 in Comatose Patients. *Audiology and Neuro-otology*, *5*, 192-197.doi: 10.1159/000013880

Fischer, C., Morlet, D., Bouchet, P., Luaute, J., Jourdan, C., &Salord, F. (1999). Mismatch negativity and late auditory evoked potentials in comatose patients. *Clinical*

Neurophysiology, 110, 1601-1610.doi: 10.1016/S1388-2457(99)00131-5

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 6, 78-84. doi:10.1016/S1364-6613(00)01839-8

Fruhstorfer, H. (1971). Habituation and dishabituation of the human vertex response. *Electroencephalography and Clinical Neurophysiology, 30*, 306-312. doi: 10.1016/0013-4694(71)90113-1

Fruhstorfer, H., Soveri, P., Järvilehto, T. (1970). Short-term habituation of the auditory evoked response in man. *Electroencephalography and Clinical Neurophysiology, 28*, 153–161. doi: 10.1016/0013-4694(70)90183-5

Gaskell, M.G., &Marslen-Wilson,W.D.(2002).Representation and competition in the perception of spoken words. *Cognitive Psychology, 45*, 220–266.doi: 10.1016/S0010-0285(02)00003-8

Gu, F., Li, J., Wang, X., Hou, Q., Huang, Y., & Chen, L. (2012). Memory traces for tonal language words revealed by auditory event-related potentials. *Psychophysiology, 49*, 1353-1360. doi: 10.1111/j.1469-8986.2012.01447.x

Harris, I.D. (1943). Habituatory response decrement in the intact organism. *Psychological Bulletin, 40*, 385-422. doi: 10.1037/h0053918

Howie, J. M. (1974). On the domain of tone in Mandarin. *Phonetica, 30*, 129-148. doi: 10.1159 /000259484

Hu, J., Gao, S., Ma, W., & Yao, D. (2012). Dissociation of tone and vowel processing in Mandarin idioms. *Psychophysiology, 49*, 1179–1190. doi:

10.1111/j.1469-8986.2012.01406.x

Huang, X., Yang, J.-C., Zhang, Q., & Guo, C. (2014). The time course of spoken word recognition in Mandarin Chinese: A unimodal ERP study. *Neuropsychologia*, *63*, 165–174. doi:10.1016/j.neuropsychologia.2014.08.015

Joanisse, M. F., Zevin, J. D., & McCandliss, B. D. (2007). Brain mechanisms implicated in the preattentive categorization of speech sounds revealed using fMRI and a short-interval habituation trial paradigm. *Cerebral Cortex*, *17*, 2084–2093. doi:10.1093/cercor/bhl124

Kandel, E.R. (2001). Neuroscience-The molecular biology of memory storage: A dialogue between gene and synapses. *Science, 294*, 1030–1038. doi: 10.1126/science.1067020

Kimppa, L., Kujala, T., Leminen, A., Vainio, M., & Shtyrov, Y. (2015). Rapid and automatic speech-specific learning mechanism in human neocortex. *NeuroImage*, *118*, 282–291. doi:10.1016/j.neuroimage.2015.05.098

Kutas, M., &Federmeier, K.D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Science, 4*, 463-470. doi: 10.1016/S1364-6613(00)01560-6

Kutas, M., Federmeier, K.D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology, 62*, 621-641. doi: 10.1146/annurev.psych.093008.131123

Kutas, M., &Hillyard, S.A., (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*, 161-163. doi: 10.1038/307161a0

Lai, Y., & Zhang, J. (2008). Mandarin lexical tone recognition: The gating paradigm.

Kansas Working Papers in Linguistics, 30, 183–194. doi: 10.17161/KWPL.1808.3914

Language Affairs Revolution Council of China. (1988). *现代汉语常用字表(List of frequently used characters in modern Chinese)*. <http://www.china-language.gov.cn/wenziguifan/shanghi/013c.htm>

Lee, C.Y. (2007). Does horse activate mother? Processing lexical tone in form priming. *Language and Speech, 50*, 101–123. doi: 10.1177/00238309070500010501

Li, N., Zhang, X.D., Huang, Z.M. (2009). 汉语鼻辅音共振峰的比较研究(A comparative study on the formants of nasal consonants in Chinese). *中国听力语言康复科学杂志 (Chinese Scientific Journal of Hearing and Speech Rehabilitation), 5*, 36-38.

MacGregor, L.J., Pulvermüller, F., van Casteren, M., & Shtyrov, Y. (2012). Ultra-rapid access to words in the brain. *Nature Communication, 3*, 711. doi: 10.1038/ncomms1715

Malins, J. G., & Joanisse, M. F. (2012a). Towards a model of tonal processing during Mandarin spoken word recognition. *Proceedings of the 3rd International Symposium on Tonal Aspects of Languages (TAL 2012)*.

Malins, J. G., & Joanisse, M. F. (2012b). Setting the tone: An ERP investigation of the influences of phonological similarity on spoken word recognition in Mandarin Chinese. *Neuropsychologia*, *50*, 2032–2043. doi:10.1016/j.neuropsychologia.2012.05.002

Mildner, V. (2008). *The Cognitive Neuroscience of Human Communication*. New York: Lawrence Erlbaum Associates.

Muenssinger, J., Stingl, K. T., Matuz, T., Binder, G., Ehehalt, S., & Preissl, H. (2013). Auditory habituation to simple tones: Reduced evidence for habituation in children

compared to adults. *Frontiers in Human Neuroscience, 7*, 377. doi: 10.3389/fnhum.2013.00377

Mulert, C., Jäger, L., Propp, S., Karch, S., Störmann, S., Pogarell, O., …Hegerl, U. (2005). Sound level dependence of the primary auditory cortex: Simultaneous measurement with 61-channel EEG and fMRI.*NeuroImage, 28*, 49–58. doi: 10.1016/j.neuroimage.2005.05.041

Näätänen, R. (1995). The mismatch negativity: A powerful tool for cognitive neuroscience. *Ear Hear, 16*, 6–18. doi: 10.1097/00003446-199502000-00002

Näätänen, R., Gaillard, A. W. K., &Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *ActaPsychologica, 42*, 313–329. doi: 10.1016/0001-6918(78)90006-9

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Houtilainen, M., livonen, A., ...Luuk, A. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385*, 432–434. doi: 10.1038/385432a0

Näätänen, R., &Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology, 24*, 375–425. doi: 10.1111/j.1469-8986.1987.tb00311.x

Nordby, H., Hammerborg, D., Roth, W. T., &Hugdahl, K. (1994). ERPs for infrequent omissions and inclusions of stimulus elements. *Psychophysiology, 31*, 544–552.doi: 10.1111/j.1469-8986.1994.tb02347.x

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory.

Neuropsychologia, 9, 97–113. doi: 10.1016/0028-3932(71)90067-4

Öhman, A., &Lader, M. (1972). Selective attention and "habituation" of the auditory averaged evoked response in humans. *Physiology and Behaviour, 8*, 79–85. doi: 10.1016/0031-9384(72)90132-1

Pang, E. W., & Taylor, M. J. (2000). Tracking the development of the N1 from age 3 to adulthood: An examination of speech and non-speech stimuli. *Clinical Neurophysiology, 111*, 388–397. doi: 10.1016/S1388-2457(99)00259-X

Pinsker, H., Kupfermann, I., Castellucci, V., &Kandel, E. (1970). Habituation and Dishabituation of the GM-Withdrawal Reflex in Aplysia. *Science, 167*, 1740–1742. doi: 10.1126/science.167.3926.1740

Prescott, J., Connolly, J. F., &Gruzelier, J. H. (1984). The augmenting/reducing phenomenon in the auditory evoked potential. *Biological Psychology, 19*, 31–44. doi: doi:10.1016/0301-0511(84)90008-5

Pulvermüller, F., Kujala, T., Shtyrov, Y., Simola, J., Tiitinen, H., Alku, P. ... Näätänen, R. (2001). Memory traces for words as revealed by the mismatch negativity. *NeuroImage, 14*, 607–616. doi: 10.1006/nimg.2001.0864

Pulvermüller, F., & Shtyrov, Y. (2006). Language outside the focus of attention: The mismatch negativity as a tool for studying higher cognitive processes. *Progress in Neurobiology, 79*, 49–71. doi: 10.1016/j.pneurobio.2006.04.004

Rankin, C. H., Abrams, T., Barry, R. J., Bhatnagar, S., Clayton, D. F., Colombo, J., … Thompson, R. F. (2009). Habituation revisited: An updated and revised description of the

behavioural characteristics of habituation. *Neurobiology of Learning and Memory, 92*, 135–138. doi: 10.1016/j.nlm.2008.09.012

Ritter, W., Vaughan, H. G., & Costa, L. D. (1968). Orienting and habituation to auditory stimuli: a study of short term changes in average evoked responses. *Electroencephalography and Clinical Neurophysiology, 25*, 550–556. doi: 10.1016/0013-4694(68)90234-4

Rosburg, T., Haueisen, J., &Kreitschmann-Andermahr, I. (2004). The dipole location shift within the auditory evoked neuromagnetic field components N100m and mismatch negativity (MMNm). *Clinical Neurophysiology, 115*, 906–913. doi: 10.1016/j.clinph.2003.11.039

Rosburg, T., Haueisen, J., & Sauer, H. (2002). Habituation of the auditory evoked field component N100m and its dependence on stimulus duration. *ClinicalNeurophysiology, 113*, 421–428. doi: 10.1016/S1388-2457(01)00727-1

Rosburg, T., Trautner, P., Boutros, N. N., Korzyukov, O. A., Schaller, C., Elger, C. E., &Kurthen, M. (2006). Habituation of auditory evoked potentials in intracranial and extracranial recordings. *Psychophysiology, 43*, 137–144. doi: 10.1111/j.1469-8986.2006.00391.x

Rosburg, T., Zimmerer, K., &Huonker, R. (2010). Short-term habituation of auditory evoked potential and neuromagnetic field components in dependence of the interstimulus interval. *Experimental Brain Research, 205*, 559–570. doi: 10.1007/s00221-010-2391-3

Rust, J. (1977). Habituation and the orienting response in the auditory cortical evoked

potential. *Psychophysiology, 14*, 123–126. doi: 10.1111/j.1469-8986.1977.tb03361.x

Sara, Y., Mozhayeva, M.G., Liu, X., Kavalali, E.T. (2002). Fast vesicle recycling supports neurotransmission during sustained stimulation at hippocampal synapses. *The Journal of Neuroscience, 22*, 1608–1617.

Schirmer, A., Tang, S.L., Penney, T.B., Gunter, T.C., Chen, H.C. (2005). Brain responses to segmentally and tonally induced semantic violations in Cantonese. *Journal of Cognitive Neuroscience, 17*, 1–12. doi: 10.1162/0898929052880057

Sharpless, S., & Jasper, H. (1956). Habituation of the arousal reaction. *Brain, 79*, 655–680. doi: 10.1093/brain/79.4.655

Shtyrov, Y. (2011). Fast mapping of novel word forms traced neurophysiologically. *Frontiers in Psychology, 2*, 340. doi: 10.3389/fpsyg.2011.00340

Shtyrov, Y. (2012). Neural bases of rapid word learning. *The Neuroscientist*, *18*, 312–319. doi:10.1177/1073858411420299

Shtyrov, Y., Butorina, A., Nikolaeva, A., &Stroganova, T. (2014). Automatic ultrarapid activation and inhibition of cortical motor systems in spoken word comprehension. *Proceedings of the National Academy of Sciences of the United States of America, 111*, E1918–E1923. doi: 10.1073/pnas.1323158111

Shtyrov, Y., Kimppa, L., Pulvermüller, F., &Kujala, T. (2011). Event-related potentials reflecting the frequency of unattended spoken words: A neuronal index of connection strength in lexical memory circuits? *NeuroImage, 55*, 658-668. doi: 10.1016/j.neuroimage.2010.12.002

Shtyrov, Y., Nikulin, V. V, & Pulvermüller, F. (2010). Rapid cortical plasticity underlying novel word learning. *The Journal of Neuroscience, 30*, 16864–16867. doi: 10.1523/JNEUROSCI.1376-10.2010

Shtyrov, Y., & Pulvermüller, F. (2002). Neurophysiological evidence of memory traces forwords in the human brain. *Neuroreport, 13*, 521-525. doi: 10.1097/00001756-200203250-00033

Shuai, L., & Gong, T. (2014). Temporal relation between top-down and bottom-up processing in lexical tone perception. *Frontiers in Behavioral Neuroscience, 8*, 1–16. doi: 10.3389/fnbeh.2014.00097

Sörös, P., Michael, N., Tollkötter, M., &Pfleiderer, B. (2006). The neurochemical basis of human cortical auditory processing: combining proton magnetic resonance spectroscopy and magnetoencephalography. *BMC Biology*, *4*, 25. doi:10.1186/1741-7007-4-25

Sörös, P., Teismann, I. K., Manemann, E., &Lütkenhöner, B. (2009). Auditory temporal processing in healthy aging: A magnetoencephalographic study. *BMC Neuroscience*, *10*, 34. doi:10.1186/1471-2202-10-34

Teismann, I. K., Sörös, P., Manemann, E., Ross, B., Pantev, C., & Knecht, S. (2004). Responsiveness to repeated speech stimuli persists in left but not right auditory cortex. *Neuroreport, 15*, 1267–1270. doi: 10.1097/01.wnr.0000129856.58404.2d

Thompson, R. F., & Spencer, W. A. (1966). Habituation: A model phenomenon for the study of neuronal substrates of behaviour. *Psychology Review, 73*, 16–43. doi: 10.1037/h0022681

Wang, H., Chang, B., Li, Y., Lin, L., Liu, J., Sun, Y., et al. (1986). *现代汉语频率词典* (Modern Chinese word frequency dictionary). Beijing: Beijing Language Institute Publisher

Wang, P., & Knösche, T. R. (2013). A realistic neural mass model of the cortex with laminar-specific connections and synaptic plasticity: Evaluation with auditory habituation. *PloS ONE, 8*, e77876. doi: 10.1371/journal.pone.0077876

Woods, D. L., &Elmasian, R. (1986). The habituation of event-related potentials to speech sounds and tones. *Electroencephalography and Clinical Neurophysiology, 65*, 447–459. doi: 10.1016/0168-5597(86)90024-9

Wu, N.N., & Shu, H. (2003). Gating技术与汉语听觉词加工 (The gating paradigm and spoken word recognition of Chinese). *心理学报 (ActaPsychologica), 35*, 582-590.

Yabe, H., Tervaniemi, M., Sinkkonen, J., Huotilainen, M., Ilmoniemi, R. J., & Näätänen, R. (1998).Temporal window of integration of auditory information in the human brain. *Psychophysiology. 35*, 615–619. doi:10.1017/s0048577298000183.

Ye, Y., & Connine, C. M. (1999). Processing spoken Chinese: The role of tone information. *Language and Cognitive Processes, 14*, 609–630. doi: 10.1080/016909699386202

Yue, J., Bastiaanse, R., & Alter, K. (2014). Cortical plasticity induced by rapid Hebbian learning of novel tonal word-forms: Evidence from mismatch negativity. *Brain and Language, 139*, 10–22. doi: 10.1016/j.bandl.2014.09.007

Zevin, J.D., &McCandliss, B.D. (2005). Dishabituation of the BOLD response to speech sounds. *Behavioral and Brain Functions, 1*, 4. doi: 10.1186/1744-9081-1-4

Zhao, J., Guo, J., Zhou, F., & Shu, H. (2011). Time course of Chinese monosyllabic spoken word recognition: Evidence from ERP analyses. *Neuropsychologia, 49*, 1761–1770. doi: 10.1016/j.neuropsychologia.2011.02.054

Appendix

Fillers used in company with the four experimental word-forms.

*/na1/: */pie2/, /che2/ (撤, “withdraw”), /fou3/ (否, “negation”)

/ma1/: */yue2/, */te3/, /nu4/ (怒, “anger”)

/ni2/: */zhuo4/, */que3/, gu1 (姑，“ante”)

/mi2/: */tou3/, */ru1/, /jue4/ (倔, “stubborn”)

Table Legend

Table 1 Spoken tonal word forms used in the experimental and control conditions

Table 2 The average N1 habituation coefficient *R* values for the four repetitions in the three ROIs (left, midline, and right), and for the four types of stimulus.

Table 3 The average peak latencies of N1 (ms) for the five presentations in the three ROIs (left, midline, and right), and for the four types of stimulus.

Figure Legend

Figure 1 Illustration of a stimulation train used for habituation design. S1 to S5 refer to the first to the fifth presentation of a particular stimulus.

Figure 2 The oscillograms (top row) and spectrograms (bottom row) of the four stimuli used. The f0 contours are illustrated with blue lines. Note the similarity between the envelopes of the two experimental stimuli (*/na1/, /ma1/) and those of the two control stimuli (/ni2/, /mi2/), and the same pitch contours of the experimental stimuli and the two control stimuli.

Figure 3 Grand averaged ERPs elicited by S1 to S5 of the four types of stimuli (*/na1/, /ma1/, /ni2/, /mi2/) at three representative electrodes from the left (C3), midline (Cz), and right (C4) scalp regions. The overlaid bars denote the decrement of N1 peak amplitudes with the repetition of stimuli. The upper bars show the N1 peaks to S1; the middle bars to the approximate peaks to the repeated stimuli; the lower bars are at the baseline. Note the greater distance between the middle bar and the upper bar at the C4 for */na1/ relative to others,

suggesting more prominent N1 decrement in the right ROI.

Figure 4 Topographic maps of the grand averaged N1 peak amplitudes for each of the presentations (S1 – S5) within the presentation train and for each of the conditions.

Figure 5 Topographic maps of the habituation coefficient *R* data collapsed across repetitions S2 to S5 by performing averaging for each of the conditions.

Figure 6 The lexicality effect of the N1 habituation. (A) Average habituation coefficient *R* data in the left, midline, and right scalp ROIs for each of the four stimulus repetitions, with separate lines for each of the four stimuli. (B) Statistical analyses on the *R* data in the right ROI for each of the four types of stimuli. The error bars show standard error of the mean (SEM). *$p < 0.05$, **$p < 0.01$, *n.s.* $p > 0.05$.

Table 1 Spoken tonal word forms used in the experimental and control conditions

	Experimental conditions		Control conditions	
Pinyin	*/na1/	/ma1/	/ni2/	/mi2/
Chinese character		妈	泥	迷
Meaning		mother	mud	mystery

Table 2 The average N1 habituation coefficient *R* values for the four repetitions in the three ROIs (left, midline, and right), and for the four types of stimulus.

	N1-S2		N1-S3		N1-S4		N1-S5	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Left ROI								
*/na1/	0.50	0.24	0.57	0.30	0.54	0.30	0.60	0.35
/ma1/	0.55	0.32	0.63	0.29	0.66	0.29	0.58	0.36
/ni2/	0.53	0.24	0.60	0.29	0.62	0.29	0.49	0.37
/mi2/	0.48	0.20	0.65	0.28	0.66	0.23	0.47	0.28
Midline ROI								
*/na1/	0.45	0.25	0.52	0.31	0.47	0.27	0.55	0.28
/ma1/	0.57	0.27	0.64	0.27	0.61	0.31	0.54	0.39
/ni2/	0.51	0.24	0.54	0.29	0.53	0.22	0.52	0.29
/mi2/	0.50	0.20	0.62	0.28	0.59	0.20	0.47	0.33
Right ROI								
*/na1/	**0.46**	0.21	**0.48**	0.28	**0.39**	0.19	**0.53**	0.30
/ma1/	**0.65**	0.22	**0.69**	0.30	**0.67**	0.28	**0.59**	0.41
/ni2/	**0.55**	0.25	**0.57**	0.34	**0.59**	0.24	**0.60**	0.31
/mi2/	**0.53**	0.25	**0.64**	0.30	**0.62**	0.22	**0.56**	0.40

Table 3 The average peak latencies of N1 (ms) for the five presentations in the three ROIs (left, midline, and right), and for the four types of stimulus.

	N1-S1		N1-S2		N1-S3		N1-S4		N1-S5	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Left ROI										
*/na1/	108	16	101	13	102	13	105	14	101	10
/ma1/	122	18	116	19	120	16	123	13	120	19
/ni2/	113	7	102	12	109	16	104	12	105	9
/mi2/	114	13	117	20	119	13	120	16	119	15
Midline ROI										
*/na1/	108	17	102	14	104	11	101	13	101	10
/ma1/	122	18	116	19	117	18	122	14	116	16
/ni2/	111	8	104	14	108	16	103	13	105	10
/mi2/	114	14	120	21	115	16	124	17	120	15
Right ROI										
*/na1/	107	16	106	15	104	11	104	14	105	17
/ma1/	122	18	116	19	117	18	114	16	117	17
/ni2/	111	9	105	12	108	16	104	12	105	11
/mi2/	111	15	117	15	115	16	120	19	119	16

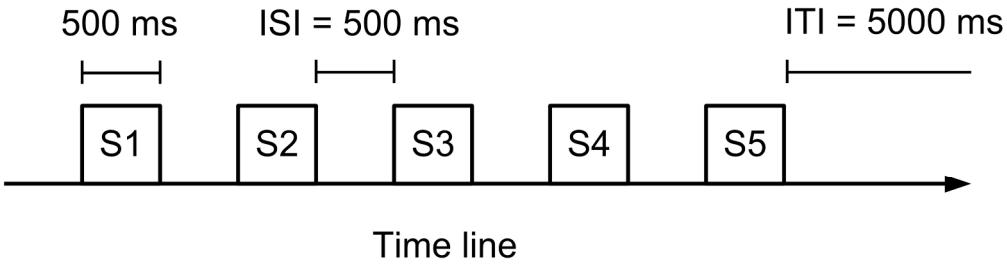


Figure 1

642x174mm (96 x 96 DPI)

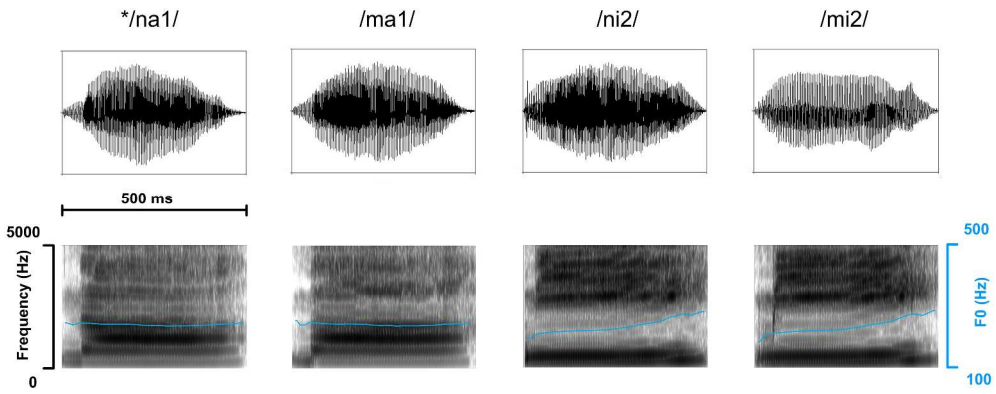


Figure 2

1032x402mm (96 x 96 DPI)

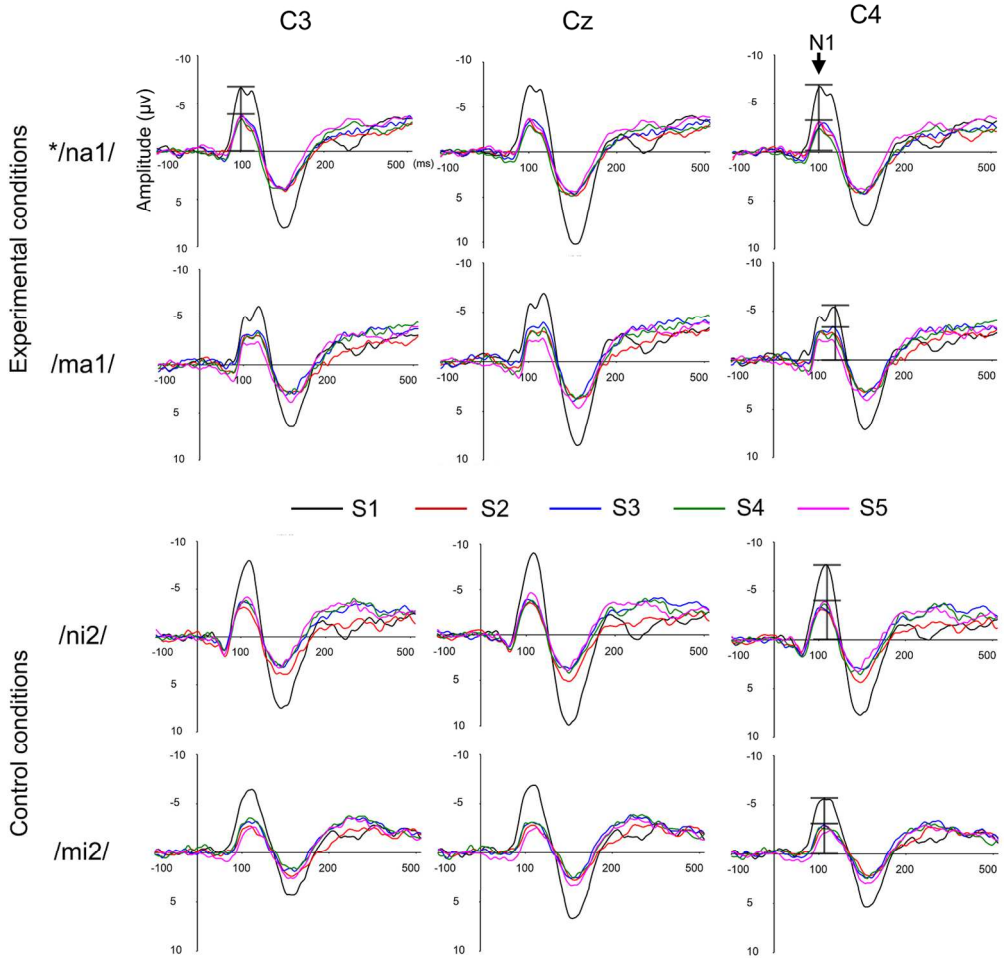


Figure 3

132x127mm (300 x 300 DPI)

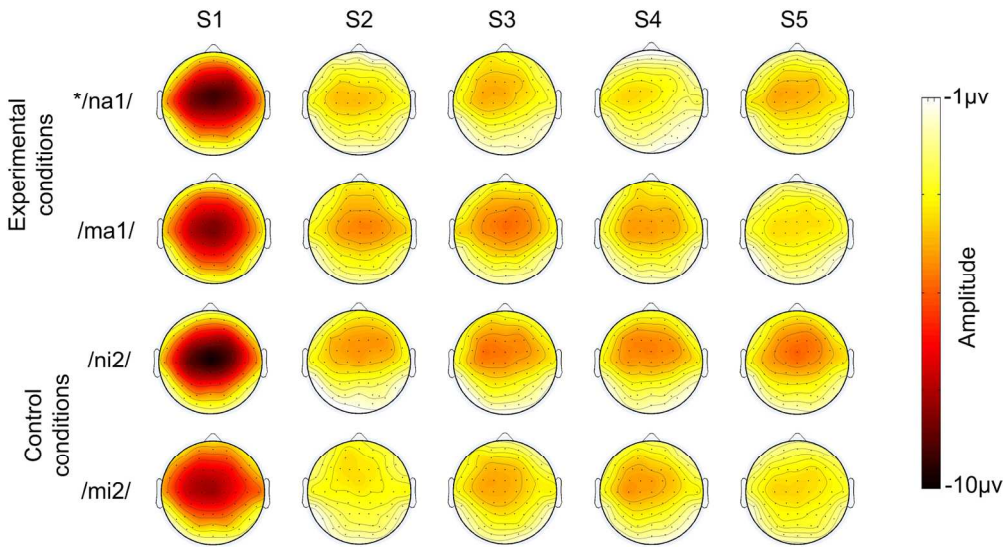


Figure 4

148x80mm (300 x 300 DPI)

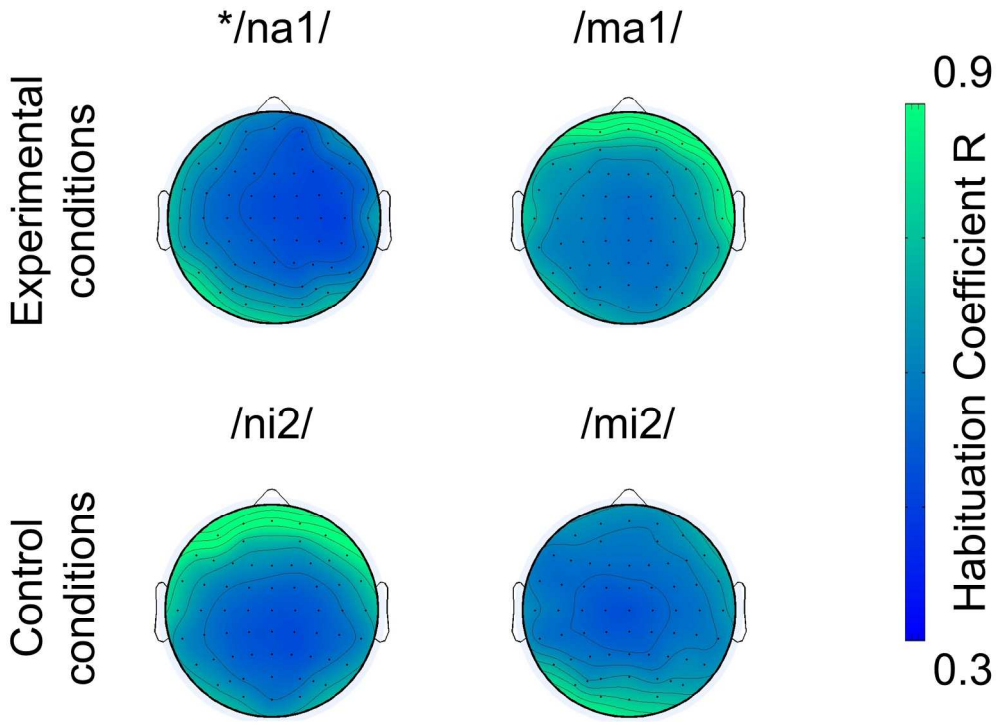


Figure 5

568x411mm (96 x 96 DPI)

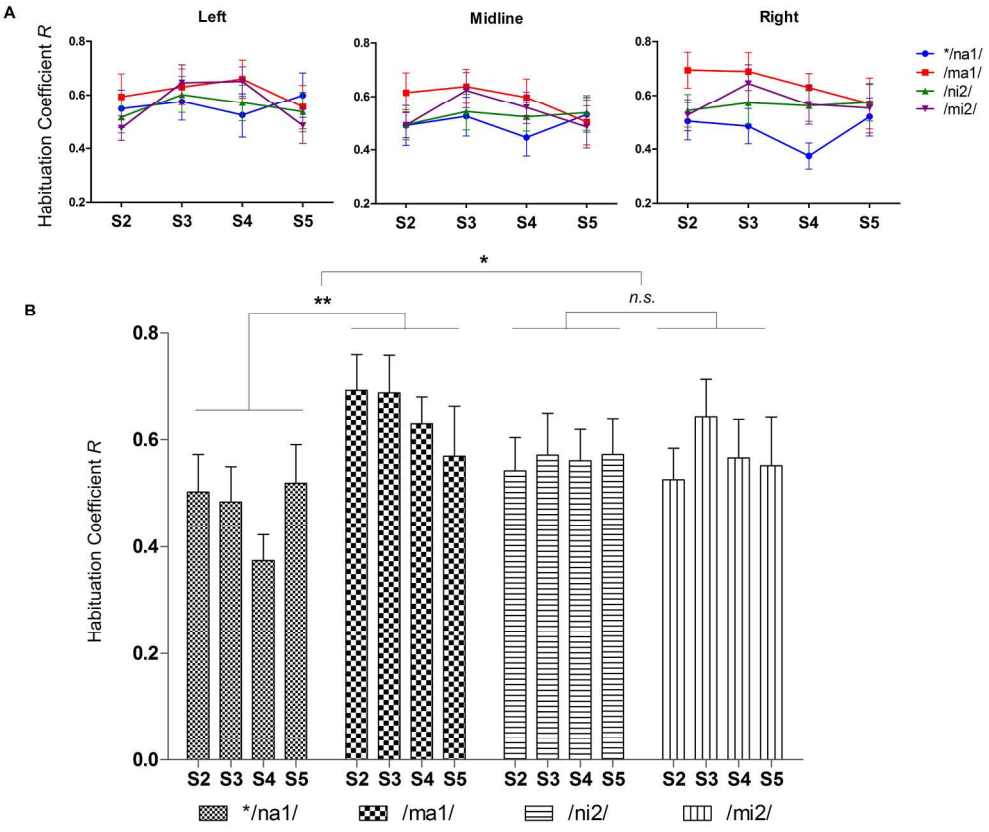


Figure 6

185x156mm (300 x 300 DPI)